

Riflessioni su problemi aperti nell'analisi degli investimenti immobiliari

Elena Fregonara¹

1. Introduzione

La teoria statistica e le pratiche della simulazione possono fornire un contributo nella risoluzione di alcuni problemi ancora aperti nell'analisi degli investimenti immobiliari: per esempio il trattamento degli effetti qualitativi, delle componenti rischiose ed incerte, delle variazioni temporali delle variabili di input nei modelli di analisi esplorati. Questi aspetti costituiscono oggetto di studio, sullo sfondo dell'analisi degli investimenti risolta secondo l'approccio probabilistico².

Nell'applicare procedure per la verifica delle attese economiche associate a progetti di investimento si ripropone il problema del reperimento, organizzazione e trattamento dei dati di mercato. Questo vale in riferimento alla sfera pubblica come a quella privata -nella consapevolezza che, alla luce delle più recenti trasformazioni nei rapporti e nelle modalità di finanziamento dei progetti, questa separazione non è più così netta come nel passato-. La formulazione dei giudizi di convenienza economica rispetto all'impiego di capitale privato da parte delle industrie immobiliari, così come la risoluzione di problemi di ingegneria economica rispetto alla destinazione di risorse pubbliche da parte dello Stato, è complicata dalla nota carenza di dati osservabili e da un contesto informativo incerto che oscura la misurazione della rischio-

1) Ricercatore universitario in Estimo presso il Politecnico di Torino, Dipartimento Casa-città.

2) Il presente scritto è la prosecuzione di uno studio teorico condotto con R. Curto. Lo studio in questione -partendo dal problema della verifica della convenienza economica in progetti di investimento nel settore immobiliare, interessati da componenti rischiose ed incerte- esplorava aspetti teorici e l'operativizzabilità della Probability Analysis quale strumento per l'analisi di rischio quantitativa, secondo gli approcci analitico e simulativo. In questa sede, proseguendo verso l'indirizzo simulativo, si riprendono alcuni nodi più specifici che, sia da un punto di vista teorico che pratico, sembrano ancora scarsamente dissodati almeno in relazione al particolare contesto immobiliare. Cfr. Curto, R.; Fregonara, E., Il controllo del rischio e dell'incertezza negli investimenti immobiliari: la probability analysis, in *Genio Rurale*, n. 9, 1997; Curto, R.; Fregonara, E., Decision tools for investments in the real estate sector with risk and uncertainty elements, in *Jahrbuch fuer Regionalwissenschaft*, n.19, 1999.

sità endogena ai progetti.

Date queste difficoltà, risulta non immediato produrre una corretta modellizzazione del sistema economico in studio, pur disponendo di tecniche deputate alla verifica della base informativa a disposizione rispetto alle specifiche esigenze analitiche, a saggiare la significatività delle informazioni disponibili e l'estendibilità dei risultati ottenuti (dal campione alla popolazione).

La statistica inferenziale -ricordando che, per dirla con Klippendorf, "l'inferenza è, evidentemente, la *raison d'être* stessa per ogni analisi del contenuto"³- da tempo si occupa di questi aspetti, mettendo a punto pratiche ampiamente studiate nei problemi di analisi dei mercati immobiliari affetti dai limiti pocanzi indicati.

Guardando al problema secondo un'altra angolazione, si prova a riflettere sull'utilizzo delle procedure di simulazione, con la convinzione che quest'ultima, esprimendo i risultati in termini di probabilità e fornendo dati sui livelli di confidenza dell'analisi, possa essere considerata quale strumento di previsione e di approssimazione di una situazione reale (Kleijnen e Van Groenendaal, 1992) e sul potenziale di analisi offerto dalla creazione di banche dati simulate quale patrimonio informativo di partenza. L'approccio simulativo, nelle sue formulazioni teoriche e pratiche, peraltro già indagato quale strumento per la risoluzione della Probability Analysis nell'analisi di rischio quantitativa, viene qui ripreso insieme alla analisi dei flussi di cassa in versione probabilistica.

Le riflessioni contenute in questo scritto, peraltro ancora frammentate e certamente suscettibili di essere completate anche con applicazioni esemplificative, nascono da alcuni spunti riconducibili a contesti disciplinari diversi: la teoria della probabilità e statistica, della quale sono esplorati strumenti da questa offerti al fine di risolvere alcuni problemi ancora aperti in procedure maturate nell'economia dell'azienda; l'econometria, dove si attingono metodi qui riletti in chiave probabilistica ed in relazione a pratiche di simulazione. Una sostanziosa e consolidata letteratura mette a punto percorsi logici ed operativi, fornendo anche una dettagliata formalizzazione matematica di concetti teorici e procedure, per quanto concerne gli aspetti di fondo; le giustificazioni teoriche dell'associazione tra gli stessi è rintracciabile invece in una trattazione più specifica e, almeno per quanto concerne la pubblicistica presa in considerazione per la stesura di questo lavoro, più estranea alla discipli-

3) Cfr. Klippendorf, K., Analisi del contenuto. Introduzione metodologica, ERI, Torino 1983, p. 84.

na estimativa.

Da un punto di vista pratico, si fa riferimento ad alcune esperienze applicative condotte con l'uso di software specifici⁴.

Un primo spunto di riflessione riguarda alcuni problemi nella modellizzazione dei sistemi economico-immobiliari, intrapresa con finalità valutative in generale e nell'analisi costi-ricavi probabilistica in particolare: problemi legati alla difficoltà nell'individuare le componenti necessarie alla rappresentazione il più oggettiva possibile del quadro reale di riferimento, dunque legate alla debole base informativa di partenza, alla presenza di componenti di incertezza e di variabili stocastiche; ancora, problemi legati alla dinamicità dei contesti modellizzati, complicati da alterazioni negli andamenti dei valori delle variabili di input nel tempo, non sempre prevedibili con precisione su base oggettiva. Queste difficoltà sono legate, più in generale, al passaggio dalla condizione ordinaria a quella reale che, nell'analisi costi-ricavi, si traduce nel problema della stima delle voci di input: come noto, le condizioni dell'ordinarietà nel caso dell'ACR non riproducono i fenomeni economici che influenzano i valori. Il riferimento è dunque allo studio intorno alla scelta della forma funzionale più adatta ad esprimere l'andamento dei valori delle variabili stocastiche in scarsità di dati osservati, ossia della funzione di distribuzione di probabilità, sia in senso statico, sia ponendosi questioni su come sia possibile internalizzare la componente temporale nell'analisi -o, meglio, la componente di incertezza associata con gli andamenti nel tempo- (Law e Kelton, 1991; Stephens e Andres, 1993; Curto e Fregonara, 1997).

Il "problema dell'input" -come possiamo sinteticamente indicare le questioni appena menzionate- è oggetto della prima parte.

Un secondo spunto di riflessione riguarda l'individuazione, interpretazione ed elaborazione dell'output dell'analisi -espresso anch'esso, coerentemente con la base informativa, in forma stocastica-. Lasciando sullo sfondo la letteratura, anche datata, che a partire dalle prime formulazioni di Hertz negli anni '60 getta i fondamenti teorici per l'analisi di rischio, si riprendono alcune più recenti riflessioni critiche sul concetto di varianza dell'output quale proxy per il rischio (Hertz, 1964;

4) Per le sperimentazioni pratiche si è ricorso all'utilizzo del software @Risk, della Palisade Corp., versione 3.1, per Windows. I risultati di alcune prime applicazioni sono rintracciabili in: Lucci, O., Le trasformazioni delle grandi aree: convenienze private e utilità pubbliche, Politecnico di Torino, Tesi di Laurea, relatore R. Curto, Anno Accademico 1995-96, I sessione, febbraio 1996.

Hughes, 1990).

Il "problema dell'output" assume un duplice significato: da un lato, riguarda le questioni legate all'analisi di sensibilità volta a cogliere gli effetti prodotti dalle variabili di ingresso incerte sui risultati ottenuti, anche in presenza di forte correlazione e alla presenza di effetti associati fra le stesse variabili di ingresso; dall'altro, si richiamano procedure econometriche ampiamente studiate quali l'analisi della varianza, la regressione multipla, o i meno noti modelli di equazioni strutturali, al fine di coglierne l'applicabilità rispetto al contesto procedurale "simulativo" e "probabilistico": si propone una reinterpretazione delle stesse in chiave stocastica, dove il tipico problema di analisi multivariata è riferito ad una base informativa costituita da dati simulati e non -o, meglio, non solo- osservati.

Dal momento, poi, che le variabili e relativi valori introdotti nel modello propagano incertezza sui risultati del modello stesso, data la soggettività endogena alle pratiche di modellizzazione e la presenza di componenti di errore -individuate e misurate attraverso procedure econometriche- ad essa conseguenti, si propone una riflessione intorno alle tecniche di "verifica" e "validazione" dei modelli proposti (in particolare sui passaggi in grado di ridurre gli effetti distorcenti prodotti sui risultati).

Infine, partendo dal presupposto che *"the perfect model would be the real system itself (by definition, any model is a simplification of the reality)"*⁵, si considerano due questioni: da un lato il confronto fra risultati ottenuti per via analitica e risultati simulati (quindi, indirettamente, il problema del campionamento); dall'altro i problemi della comparazione fra dati reali e dati simulati.

In aggiunta alle procedure più consolidate per le operazioni di verifica e validazione necessarie, si propone di sperimentare i modelli di equazioni strutturali in qualità di tecniche per la riduzione della varianza dei risultati, attraverso la ripulitura dei risultati medesimi dalle componenti di errore. Già alcune varianti alla regressione multipla, si menziona, sono state esplorate al fine di rendere minima la varianza ottenuta con applicazioni basate sull'utilizzo di stimatori non condizionati, di ridurre l'intervallo di fiducia, di individuare e scindere le componenti dell'errore e di aumentare in generale la capacità previsiva dei modelli; in questo

5) Kleijnen, J.P.C., Verification and validation of simulation models, in European Journal of Operational Research, n. 82, 1995, p. 145.

senso, si propone nuovamente il MES in qualità di alternativa all'uso di stimatori condizionati.

2. Il "problema dell'input"

Se, come afferma Helton, "*it is assumed that the model under consideration can be represented by a function of the form*

$$y = f(x_1, x_2, \dots, x_n) = f(x) \quad (1)$$

where y is the model prediction (i.e. the output variable) and x_1, x_2, \dots, x_n are the model inputs (i.e. the input variables)"⁶, con l'espressione "problema dell'input" facciamo riferimento alla individuazione delle variabili indipendenti di questa funzione.

Qualora, come nel nostro caso, oggetto della modellizzazione è un sistema economico-immobiliare e qualora ci affidiamo ai risultati di una analisi dei flussi di cassa in termini probabilistici -risolta mediante simulazione- per decidere sull'opportunità di un investimento, si parte da alcune assunzioni:

- si dispone di una adeguata base informativa;
- si dispone di soglie di accettabilità delle quote di rischio e di redditività, in segmenti di mercato comparabili, associate all'investimento in oggetto.

Per base informativa adeguata intendiamo una collezione di dati organizzati e temporalmente riferiti, relativi a prezzi e valori "certi" su costi e ricavi del progetto da valutare e, inoltre, di sufficienti dati (parametri) per la definizione delle *Probability Density Functions* delle variabili "critiche"⁷.

Mentre nel secondo caso non può essere elusa la condizione di fare

6) Si veda in proposito il saggio Helton, J.C., "Uncertainty and sensitivity analysis techniques for use in performance assessment for radioactive waste disposal", in *Reliability Engineering and System Safety*, 42, pp. 327-367, 1993. L'espressione è riportata a pg. 328.

7) Concetti teorici sulla costruzione delle "PDFs" sono rintracciabili in: Stephens, M.E.; Goodwin, B.W.; Andres, T.H., *Deriving parameter probability density functions*, in *Reliability Engineering and System Safety*, 42, pp. 271-291, 1993.

riferimento a dati di mercato, nel primo caso -partendo dalla convinzione che la simulazione può migliorare la base informativa, quindi la decisione- si può fare ricorso a dati simulati, a partire comunque dalla disponibilità di osservazioni di mercato in base alle quali individuare i parametri delle funzioni di densità di probabilità. Hughes (1990) afferma che la ricostruzione dei sistemi di prezzi e valori con finalità valutative e previsive può essere ottenuta mediante simulazione la quale, introducendo elementi di oggettività attraverso le distribuzioni dei valori, prevede sistemi appunto di prezzi e valori che siano compatibili con la struttura periodale del cash-flow e la natura probabilistica dei valori di input. Inoltre le funzioni random possono essere usate come soluzione per le deficienze dei dati, superando i problemi provocati dall'uso di proxy che condizionano i risultati, o al ricorso a modelli di costruzione dei prezzi condotti in modo erroneo⁸.

Nella "costruzione" della base informativa per l'analisi una questione di fondo è la simultaneità con cui si verificano e devono essere trattati i diversi aspetti dell'incertezza la quale, dato il carattere previsivo dell'analisi, presenta già di per sé un andamento crescente con il tempo, associata ad una componente di errore non ignorabile; inoltre, i dati di cui si dispone dovrebbero essere in grado di riprodurre le fluttuazioni del mercato. Questi aspetti devono essere considerati nella fase di produzione delle curve di distribuzione probabilistiche (da cui dipenderanno i diversi scenari), sulla base del fitting dei dati in possesso. Le questioni da affrontare sono dunque:

- la definizione delle singole variazioni e delle variazioni congiunte delle variabili critiche di ingresso;
- la definizione degli andamenti e delle fluttuazioni dei valori nel tempo e la conseguente previsione delle variazioni periodali lungo le "linee di flusso".

Quest'ultimo aspetto, senza pretese di esaustività, è oggetto del sottoparagrafo che segue.

2.1. *L'internalizzazione della componente temporale nel quadro crono-finanziario*

L'analisi costi-ricavi (Prizzon, 1995) si fonda da un lato su un problema di previsione, dall'altro di definizione dei valori di input al

8) Si veda Hughes, W.T., Valuation and risk analysis of lease properties through simulation: an objective Monte Carlo Technique, dissertazione di dottorato, Università della Georgia, 1990.

quadro crono-finanziario (o, impropriamente, modello). Come si è già detto, tutti gli elementi che compongono il riferimento dell'analisi sono affetti da una condizione di incertezza crescente con il tempo e intrinseca alla natura previsiva dell'analisi stessa.

Questo aspetto di fondo è reso più complesso dallo sfasamento temporale rispetto al ciclo produttivo fra costi e ricavi: in teoria, una più esatta stima di questi ultimi sarebbe determinabile solo alla fine della vita del progetto ma, nella prassi, si è soliti procedere considerando i costi all'inizio del ciclo produttivo ed i ricavi alla fine del ciclo medesimo. La misura dei costi -ossia l'acquisizione dei fattori produttivi o in altre parole la misura dell'investimento- e dei ricavi -ossia la misura del disinvestimento derivante dalla vendita dei prodotti è rappresentata dalla corrispondente variazione numeraria.

Un problema cruciale è costituito dalla necessità di internalizzare le variazioni temporali che interessano le variabili di ingresso: variazioni legate agli andamenti dei sistemi economici di riferimento, a fluttuazioni casuali registrate in componenti esterne o interne ai sottomercati specifici, etc. Come osservano Kleijnen e Van Groenendaal, questa difficoltà è legata più in generale al fatto che i sistemi economici non assumono il tempo come continuo: nello schema dei flussi di cassa, i diversi periodi -le celle del foglio elettronico- discretizzano la funzione temporale; il VAN è calcolato applicando una funzione nella quale il tempo compare come elemento discreto. Sintetizzando il ragionamento degli autori⁹ e richiamando la formula per il calcolo del flusso di cassa attualizzato:

$$VAN = \sum_{t=0}^T FC_t / (1 + r)^t , \quad (2)$$

si nota che i flussi di cassa FC sono considerati nei differenti periodi t, per tutto l'orizzonte temporale T, con $t = 0, 1, \dots, T$. Inoltre, i flussi di cassa sono funzioni non lineari di n variabili di input x_i -nel nostro caso "Costi di costruzione", "Oneri di urbanizzazione", "Ricavi da vendite", etc.-¹⁰; la formula (1) diventa dunque:

9) Si veda la formalizzazione proposta in Kleijnen, J.P.C.; Van Groenendaal, W., *Simulation, a statistical perspective*, Wiley, 1992, pp. 75-78. Le espressioni (2) e (3) sono rintracciabili alle pagine 75 e 76 del testo.

10) Si fa riferimento ad una Tabella dei flussi di cassa relativa ad un ipotetico progetto

$$FC_t = f(x_{1t}, x_{2t}, \dots, x_{it}, \dots, x_{nt}) \quad \text{con } t = 0, 1, \dots, T. \quad (3)$$

Le singole x_{it} sono esprimibili attraverso le loro specifiche distribuzioni (se le variabili sono intese in termini stocastici, assumendo in tal caso la forma x_{it}): in questo caso le distribuzioni, Fit , devono essere note a priori. Le variabili possono anche essere relazionate fra loro: per esempio, se x_{1t} rappresenta il prezzo di vendita unitario e x_{2t} rappresenta il volume delle vendite nel tempo, il prodotto delle due quantità fornisce il ricavo ottenuto dalle vendite nell'arco temporale preposto.

Dunque, fissato un certo tasso di sconto, l'espressione del VAN (in questo caso inteso nella forma stocastica, ossia VAN) attraverso il calcolo dei flussi di cassa stocastici FC_t internalizza la componente temporale e l'incertezza nel modello.

Al fine di esemplificare quanto esposto, nelle Tabelle 1a e 1b è riportata una descrizione sintetica delle variabili di ingresso di una analisi dei flussi di cassa applicata ad un ipotetico progetto di investimento¹¹.

Come si nota, le variabili sono identificate a mezzo del nome e dell'unità di misura; sono poi distinte per tipo (variabili deterministiche, stocastiche statiche nel tempo, stocastiche dinamiche nel tempo). Nella colonna "valore" viene riportato il valore fisso per le variabili deterministiche ed il valore atteso ricavato dalla funzione di distribuzione che definisce la variabile (che compare peraltro nel foglio elettronico sul quale è applicata la simulazione), funzione indicata nella colonna successiva e definita a mezzo dei rispettivi parametri nella colonna "Range"; successivamente, sono specificate la fonte dei dati, il comportamento rispetto al flusso di cassa (se determinano quantità monetarie in entrata o in uscita) ed infine, applicando una opportuna analisi (per esempio una matrice di correlazione), i valori di correlazione per ciascuna variabile.

La presenza della componente temporale nelle applicazioni operative inoltre porta, come sottolineano gli autori, a distinguere tra il

di investimento per la realizzazione di edilizia residenziale privata. Un esempio con variabili di input stocastiche è rintracciabile in Fregonara, E., *Forecasting and control of market risk in real estate investments*, in *Proceedings della IX Conferenza annuale della Society for Risk Analysis - SRA-Europe*, Rotterdam, The Netherlands, 10-13 ottobre 1999.

11) Le Tabelle fanno riferimento ad un quadro crono-finanziario del tipo presentato in: Fregonara, E., *Forecasting...*, op.cit.

Tabella 1a: Variabili di Input in una ACR: il caso di un progetto di investimento nel segmento immobiliare residenziale.

Variable	Un.diMis	Tipo	Valore	Distribuzione	Range	Fonte	Use/Entr	Corr/Nonco
TERRENO	L/mq	stoc/stat	valore atteso	FDD	(interv. valori)	dati di merc.	uscita	da calcolare
SPESETEC	%	det	valore	-	-	fonti uff.	uscita	da calcolare
SPESEGEN	%	det	valore	-	-	fonti uff.	uscita	da calcolare
MARKETING	%	det/din	valore	-	-	fonti uff.	uscita	da calcolare
DEMOLIZ	L/mc-mq	det	valore	-	-	dati di merc.	uscita	da calcolare
ONERIURBAN	L/mc	det	valore	-	-	Comuni	uscita	da calcolare
CONSTRSTRPAR	L/mq	det	valore	-	-	dati di merc.	uscita	da calcolare
COSTRRESID	L/mq	stoc/din	valore atteso	FDD	(interv. valori)	dati di merc.	uscita	da calcolare
COSTRBOX	L/mq	stoc/stat	valore atteso	FDD	(interv. valori)	dati di merc.	uscita	da calcolare
COSTRVERDE	L/mq	stoc/stat	valore atteso	FDD	(interv. valori)	dati di merc.	uscita	da calcolare
VENDITAAPP	L/mq	stoc/din	valore atteso	FDD	(interv. valori)	dati di merc.	entrata	da calcolare
INTERATT	%	stoc/din	valore atteso	FDD	(interv. valori)	dati di merc.	entrata	da calcolare
INTERPASS	%	stoc/din	valore atteso	FDD	(interv. valori)	dati di merc.	uscita	da calcolare
TASS.PROFIT	%	stoc/din	valore atteso	FDD	(interv. valori)	fonti uff.	uscita	da calcolare
SAGGIOSCONT	%	stoc/din	valore atteso	FDD	(interv. valori)	fonti uff.	uscita	da calcolare
INFLAZ	%	stoc/din*		da calcolare	(interv. valori)	fonti uff.	uscita	da calcolare
PERCDIVEND	%	stoc/din**		«Macro»	(interv. valori)	dati di merc.	entrata	da calcolare

* per progetti con orizzonti temporali di breve-medio termine, può ritenersi costante. Per orizzonti temporali più dilatati, si può per esempio calcolare l'andamento relativo ad un adeguato numero di anni.

** si intende il valore ottenuto dalla media delle percentuali di quantità vendute nei diversi periodi del piano-vendite.

Tabella 1b: Variabili di input in una ACR: esempio.

Variable	Un.diMis	Tipo	Valore	Distribuzione	Range	Fonte	Use/Entr	Corr/Nonco
TERRENO	L/mq	stoc/stat	valore atteso	Triangolare	(min, med,max)	dati di merc.	uscita	da calcolare
SPESETEC	%	det	5,00	-	-	fonti uff.	uscita	da calcolare
SPESEGEN	%	det	4,00	-	-	fonti uff.	uscita	da calcolare
MARKETING	%	det/din	2,00	-	-	fonti uff.	uscita	da calcolare
DEMOLIZ	L/mc-mq	det	25.000	-	-	dati di merc.	uscita	da calcolare
ONERIURBAN	L/mc	det	70.000	-	-	Comuni	uscita	da calcolare
CONSTRSTRPAR	L/mq	det	100.000	-	-	dati di merc.	uscita	da calcolare
COSTRRESID	L/mq	stoc/din	valore atteso	Triangolare	(min, med,max)	dati di merc.	uscita	da calcolare
COSTRBOX	L/mq	stoc/stat	valore atteso	Triangolare	(min, med,max)	dati di merc.	uscita	da calcolare
COSTRVERDE	L/mq	stoc/stat	valore atteso	Triangolare	(min, med,max)	dati di merc.	uscita	da calcolare
VENDITAAPP	L/mq	stoc/din	valore atteso	Normale	(med, ds)	dati di merc.	entrata	da calcolare
INTERATT	%	stoc/din	valore atteso	Uniforme	(min, max)	dati di merc.	entrata	da calcolare
INTERPASS	%	stoc/din	valore atteso	Uniforme	(min, max)	dati di merc.	uscita	da calcolare
TASS.PROFIT	%	stoc/din	valore atteso	Uniforme	(min, max)	fonti uff.	uscita	da calcolare
SAGGIOSCONT	%	stoc/din	valore atteso	Uniforme	(min, max)	fonti uff.	uscita	da calcolare
INFLAZ	%	stoc/din*		da calcolare	(interv. valori)	fonti uff.	uscita	da calcolare
PERCDIVEND	%	stoc/din**		«Macro»	(interv. valori)	dati di merc.	entrata	da calcolare

Metodo Monte Carlo, basato sull'utilizzo di numeri casuali sottostanti a precise condizioni e la simulazione, basata sulla risoluzione numerica di modelli dinamici (cioè modelli che includono una dimensione temporale)¹².

E' altresì utile menzionare, a proposito della fase di "model sampling" (la quale ricordiamo costituisce, insieme all'"iteration", uno dei passaggi su cui fonda la procedura), che il primo momento ossia la generazione di random numbers procede con la produzione di sequenze di numeri aleatori, ricorrendo a metodi aritmetici o numerici: obiettivo è la generazione di percorsi temporali alternativi nel tempo i quali, anche se possono di fatto essere singolarmente organizzati su una base deterministica, introducono elementi di dinamicità temporale all'interno dell'analisi.

Tra le voci di input critiche in un quadro dei flussi di cassa, alcune si prestano immediatamente per esemplificare il problema: il tasso di inflazione, che può essere ipotizzato crescente con il tempo e con le variazioni dei sovrasistemi economici di riferimento; il saggio di sconto, variabile che come è noto può tenere conto del valore temporale del denaro e del rischio del flusso di cassa (componente che deve però essere omessa qualora si proceda con una analisi di rischio), la quale è considerata nelle analisi sperimentate come variabile deterministica ma che potrebbe essere, per esempio, trattata attribuendo un andamento di crescita di tipo esponenziale. O ancora -aspetto su cui riponiamo particolare interesse- variazioni legate ad andamenti fluttuanti nelle vendite: i piani vendita, riproducendo i relativi sottomercati, organizzano temporalmente le percentuali di prodotto venduto incrociate con le "relative percentuali di rate pagate, ammettendo però un solo andamento possibile per ogni schema proposto.

Il problema della previsione in una analisi costi-ricavi si fa infatti più complesso nel piano vendite che in altre parti, dal momento che dipende in maniera meno diretta da scelte operate dall'investitore e in misura maggiore da decisioni derivanti dall'esterno, in particolare dalle risposte date dal mercato. In altri termini, la questione consiste nella trattazione di funzioni probabilistiche riferite a periodi temporali diver-

12) Cfr. Kleijnen, J.P.C.; Van Groenendaal, W., *Simulation* ..., op.cit., pg. 12: "We speak of Monte Carlo method whenever the solution makes use of random numbers, which are uniformly and independently distributed over the interval from zero to one. We speak of simulation whenever the model has a time dimension (the model is called dynamic), and it is solved numerically".

si, che possono essere fra loro concatenati per mezzo di legami consequenziali¹³: è corretto ritenere, infatti, che le percentuali compravvendute cambiano nel tempo e gli andamenti delle vendite nei diversi periodi sono influenzate in diverso modo dalle quantità vendute nei periodi precedenti (effetto di politiche di vendita, fattori di natura irrazionale, etc.).

Se poi si assumono i principi della probabilità condizionata menzionando al contempo il calcolo delle probabilità congiunte, è possibile trattare i ricavi tenendo conto della loro dimensione temporale, sia pure in presenza di difficoltà legate al fatto che i fenomeni in evoluzione nel tempo seguono "leggi non compiutamente conosciute o comunque non di tipo deterministico"¹⁴.

A partire da tali premesse, il problema temporale dei piani vendita può essere risolto, a nostro avviso, secondo due indirizzi:

- 2) inserendo negli appositi programmi di calcolo funzioni di distribuzione per ciascun intervallo periodale, "condizionate" in modo tale che la sommatoria dei valori campionati rispetti appunto precise condizioni (per esempio che non superi, come è triviale pensare, il 100% delle vendite)¹⁵;
- 3) passando dal concetto di variabile al concetto di processo stocastico, ricorrendo alla teoria che tratta dei "moti Browniani" (o altri processi di diffusione più generali) in grado di operare con fenomeni che seguono un andamento deterministico a cui è aggiunta una componente aleatoria¹⁶.

13) Cfr. Palisade Corp., @Risk. Advanced risk analysis for spreadsheets, User's Guide, 1995, in particolare la sezione relativa all'uso delle funzioni "Macro".

14) Cfr. Curto, R.; Fregonara, E., *Il controllo...*, op. cit., p. 30.

15) Questo primo indirizzo è stato sviluppato nell'ambito di una Tesi di Laurea presso il Politecnico di Torino, la quale costituisce un primo step nello studio del problema del trattamento della variabile temporale nel piano vendite di una analisi costi-ricavi; per le sperimentazioni applicative è stato utilizzato il software @Risk, prima menzionato. I risultati delle applicazioni e le riflessioni teoriche sono rintracciabili in: Ferraris, I., Valutazione degli investimenti immobiliari e rischio, Tesi di Laurea, relatore R. Curto, Anno Accademico 1996-97, II sessione, luglio 1997.

16) Hughes, W.T., Valuation..., op. cit., p. 39 e seguenti. L'autore esamina il moto Browniano nelle versioni aritmetica e geometrica -Arithmetic Brownian motion (ABM) e Geometric Brownian motion (GBM)- in relazione al mercato dell'affitto; a partire dalle proposte di Kassouf, l'autore indaga aspetti teorici e procedurali sostenendo che "Moto Browniano Geometrico e tecniche di simulazione Monte Carlo possono essere combinate per modellizzare il mercato dell'affitto e fornire una analisi oggettiva delle proprietà all'interno di questo mercato" (p. 49; traduzione di E.Fregonara).

Nel primo caso, il software utilizzato presenta la possibilità di inserire i valori singoli o le distribuzioni con riferimento a singole celle e, laddove necessario, di inserire relazioni di collegamento fra le celle singole, riproducendo andamenti predeterminati; inoltre permette di "condizionare" il campionamento dei valori dalle distribuzioni di probabilità ricorrendo a funzioni logiche di verifica, peraltro proprie dei linguaggi di programmazione. E' possibile infatti associare alle fasi di sampling e iterazione funzioni macro appositamente costruite.

Nel secondo caso, invece, la linea del tempo viene considerata come un vero e proprio processo stocastico. Questa interpretazione ben si adatta ad affrontare quelle analisi in cui i valori sono riferiti a momenti temporali diversi: le fluttuazioni dei valori dei beni nel tempo, i relativi prezzi, le variazioni nei tassi, etc., ed in particolare il piano-vendite dell'analisi costi-ricavi che pare appunto riprodurre una situazione dove, nella generalità dei casi, è distinguibile una componente aleatoria associata ad un andamento deterministico.

A nostro parere, può essere utile indirizzare la ricerca in entrambe le direzioni: nel primo caso, enfatizzando aspetti tecnici ed esplorando le potenzialità dei software (in particolar modo la possibilità di creare, mediante adatti linguaggi di programmazione, funzioni per la personalizzazione del calcolo secondo gli specifici problemi di analisi); nel secondo caso -muovendosi su un terreno ancora scarsamente dissodato- si tratta di considerare i legami fra processi stocastici (cioè le funzioni in grado di modellarli) e simulazione, nonchè di rivedere la base analitica di partenza in vista di una maggiore coerenza temporale e oggettività del modello.

3. Il "problema dell'output"

In questa seconda parte dello scritto sono riprese alcune tecniche econometriche -analisi della varianza, regressione multipla, modelli di equazioni strutturali- in luce diversa da altri contesti di analisi in cui sono state ampiamente studiate. Si parte da un preciso presupposto: quando, attraverso una procedura di simulazione, riusciamo a disporre di output analitici, ossia per ciascun valore della variabile di output selezionata sono noti i valori degli input, disponiamo di un patrimonio informativo che ci consente di spiegare l'interazione fra le variabili.

Le analisi di sensibilità e degli scenari, come è noto, costituiscono

una via nella risoluzione dell'analisi di incertezza: per la loro applicabilità (secondo l'approccio seguito nei programmi di calcolo da noi utilizzati) è infatti necessario disporre di collezioni di sufficienti dati campionati - cosa peraltro resa possibile dalle funzioni di memorizzazione su archivi di dati temporanei provenienti dai valori casualmente estratti durante i processi di simulazione (in Figura 1 è riportato uno stralcio esemplificativo di dati provenienti da campionamento casuale¹⁷). In questa logica operativa, all'uso di campioni di dati osservati come base informativa per l'applicazione delle procedure econometriche, viene sostituito l'uso di archivi di dati generati dal campionamento, verificati e validati a mezzo di opportuni passaggi. La questione, se ragioniamo in questi termini, si presenta dunque come un problema di analisi multivariata, in cui le variabili dipendenti possono essere per esempio gli indicatori sintetici di redditività (Valore Attuale Netto e Tasso Interno di Rendimento), pensando le quantità di input dell'analisi costi-ricavi quali variabili esplicative. Per esempio, ammesso che sia possibile l'introduzione dei ricavi sotto forma di variabile stocastica ("condizionata") inglobando così l'incertezza temporale delle vendite, è possibile tradurre i singoli piani vendita in variabili che rappresentano il "tempo di vendita".

Nei punti che seguono proviamo a trasferire sul piano simulativo alcune riflessioni intorno a limiti e potenzialità delle procedure econometriche sperimentate in altri contesti.

Come è noto, l'analisi della varianza è una tecnica che -seppure fondata su una sostanziosa base statistica- presenta finalità prevalentemente descrittive piuttosto che previsive. Certi contesti di analisi particolarmente consoni alle potenzialità della tecnica hanno peraltro messo in luce come i "limiti" rilevati a proposito della natura delle variabili trattabili siano compensati dalla capacità della tecnica stessa di elaborare variabili categoriali e qualitative, oltre che di scomporre l'output in effetti principali e secondari.

Per altri versi, la regressione multipla nelle sue varie declinazioni, comunque vincolate al trattamento delle variabili qualitative, è stata

17) L'esempio riporta i dati ottenuti dall'applicazione del programma @Risk sulla Tabella ACR rintracciabile in: Fregonara, E., *Forecasting and control of market risk in real estate investments*, in *Proceedings della IX Conferenza annuale SRA-E, Society of Risk Analysis - Europe*, Rotterdam, ottobre 1999. In particolare, nell'esempio qui esposto si è proceduto con una simulazione mediante campionamento *Latin Hypercube Sampling* (LHS), per 400 iterazioni. Altre applicazioni hanno portato anche a collezioni di 10.000 dati, la cui significatività risulta peraltro garantita da funzioni di "monitoraggio della convergenza" e di *auto-stop* della fase di *iteration* quando i dati non superano soglie percentuali minime precedentemente fissate.

Figura 1 Data-base composto da collezione di dati di simulazione: campionamento LHS, 400 iterazioni (stralcio).

345	L 633.037.200	11,34%	L 1.167.290.000	...	L 812.179.800	25735,9	1226721	3459793	2,76%	7,44%	3,70%	4,70%
346	L 858.098.700	13,98%	L 1.089.874.000	...	L 852.472.400	238739,3	1287755	3825945	2,79%	8,44%	5,04%	3,19%
347	L 439.560.700	9,45%	L 1.113.325.000	...	L 734.279.200	244081,5	1086464	3135814	2,25%	6,55%	2,18%	3,78%
348	L 760.827.200	15,65%	L 995.841.300	...	L 616.402.800	2177555	1258571	3475254	2,77%	7,77%	4,25%	3,45%
349	L 475.688.700	9,59%	L 1.153.121.000	...	L 790.364.000	238042,4	1213717	3850318	2,29%	6,74%	3,68%	3,74%
350	L 1.503.188.000	21,12%	L 1.121.443.000	...	L 863.853.000	2340092	1214238	3853338	2,08%	7,38%	5,15%	2,44%
351	L 1.256.887.000	18,89%	L 1.051.491.000	...	L 867.771.800	266042	1263701	3288966	2,08%	6,30%	3,41%	4,20%
352	L 123.635.300	1,63%	L 927.424.000	...	L 768.453.800	202678,8	1131720	3332688	2,05%	6,28%	2,33%	2,11%
353	L 1.336.890.000	16,55%	L 1.027.443.000	...	L 781.852.400	244326,6	1228619	3332688	2,05%	6,28%	2,33%	2,11%
354	L 1.336.890.000	16,55%	L 1.027.443.000	...	L 551.172.100	234526,6	1228619	2418982	2,05%	6,28%	2,33%	2,11%
355	L 1.336.890.000	16,55%	L 1.027.443.000	...	L 880.206.500	238558,8	1134729	3730122	2,51%	7,96%	5,03%	5,37%
356	L 828.152.100	13,69%	L 1.050.540.000	...	L 802.224.400	229456,8	1356907	3675753	2,46%	7,80%	5,35%	5,37%
357	L 315.710.400	7,73%	L 1.164.516.000	...	L 748.571.300	284168	1123851	3675753	2,42%	7,80%	5,35%	5,37%
358	L 1.150.304.000	17,01%	L 1.164.516.000	...	L 874.535.200	255164,8	1208223	3711580	2,40%	7,41%	3,23%	2,88%
359	L 1.381.886.000	20,91%	L 987.489.200	...	L 860.162.500	215951	1158014	3850425	2,40%	7,41%	3,86%	2,39%
360	L 263.784.800	7,11%	L 1.156.936.000	...	L 783.110.100	252328,4	1307078	3344683	3,00%	6,50%	2,90%	4,57%
361	L 1.192.801.000	18,46%	L 1.064.095.000	...	L 828.880.800	232098,9	1107552	3508158	2,86%	6,43%	5,79%	5,04%
362	L 277.688.900	7,16%	L 1.286.205.000	...	L 776.853.400	287166,4	1224801	3314587	2,85%	6,19%	5,15%	4,80%
363	L 291.330.100	0,35%	L 1.075.981.000	...	L 704.418.200	232521,6	1314008	3031411	2,15%	6,71%	4,18%	2,41%
364	L 322.784.700	0,02%	L 1.083.823.000	...	L 717.131.000	238851,4	1338315	3083313	2,81%	7,78%	5,91%	3,94%
365	L 606.892.300	11,25%	L 1.124.833.000	...	L 793.136.000	244312,4	1222373	3392560	2,50%	6,21%	4,51%	2,19%
366	L 1.576.428.000	20,75%	L 1.302.613.000	...	L 925.475.500	268096,8	1128400	3813278	2,86%	7,29%	2,39%	3,53%
367	L 794.069.400	13,13%	L 1.228.629.000	...	L 808.079.400	268573,8	1140798	3444876	2,12%	6,00%	4,78%	3,72%
368	L 424.261.800	9,23%	L 1.063.755.000	...	L 792.592.400	232789,8	1210639	3258620	2,28%	7,11%	3,57%	2,67%
369	L 208.516.300	6,74%	L 932.905.000	...	L 717.124.800	203868,8	1210639	3074319	2,08%	6,83%	4,19%	4,68%
370	L 1.092.143.000	15,55%	L 1.302.241.000	...	L 800.988.800	229062	1219713	3694256	2,08%	6,19%	3,53%	3,88%
371	L 1.055.151.000	16,86%	L 1.016.100.000	...	L 830.988.800	229062	1231018	3694256	2,08%	6,19%	3,53%	3,88%
372	L 415.658.000	8,61%	L 1.358.525.000	...	L 808.950.700	270125,6	1268017	3441573	2,37%	6,75%	3,31%	4,77%
373	L 1.083.581.000	19,54%	L 1.149.261.000	...	L 897.054.500	251812,2	1184489	3888529	2,30%	6,97%	3,94%	4,53%
374	L 1.083.581.000	19,54%	L 1.149.261.000	...	L 738.909.800	258980,4	1184489	3888529	2,30%	6,97%	3,94%	4,53%
375	L 824.574.000	23,57%	L 1.229.634.000	...	L 956.803.000	287788,1	1212848	3153748	2,70%	7,48%	5,17%	2,82%
376	L 1.380.448.000	20,54%	L 1.024.917.000	...	L 728.768.800	223251,5	1150153	4043814	2,57%	7,85%	3,14%	2,62%
377	L 99.676.220	0,01%	L 1.054.917.000	...	L 669.857.700	230848,7	1284737	3122975	2,46%	6,45%	4,48%	2,31%
378	L 1.427.191.000	20,59%	L 1.171.107.000	...	L 848.520.600	257115,9	1155559	3802424	2,78%	6,96%	2,02%	4,88%
379	L 64.632.780	4,82%	L 1.016.100.000	...	L 735.928.300	222011,3	1021657	3591658	2,51%	6,34%	3,39%	4,74%
380	L 1.419.132.000	19,03%	L 1.327.548.000	...	L 919.499.300	2291449,2	1275839	3154988	2,43%	7,02%	5,69%	2,40%
381	L 425.026.200	9,09%	L 1.174.982.000	...	L 757.924.800	257609,8	1188002	3895758	2,57%	6,15%	5,09%	2,40%
382	L 855.977.200	13,62%	L 1.203.345.000	...	L 849.788.300	287545,8	1163252	3240507	2,07%	6,06%	2,60%	4,72%
383	L 1.311.482.000	19,12%	L 1.113.161.000	...	L 887.423.400	243988,1	1224263	3808552	2,97%	8,03%	4,57%	3,32%
384	L 1.151.348.000	17,09%	L 1.196.180.000	...	L 849.747.600	262428,1	1230243	3761833	2,73%	6,42%	2,04%	2,60%
385	L 955.872.900	14,75%	L 1.243.538.000	...	L 837.325.700	272933,7	1121384	3810947	2,28%	7,28%	5,55%	5,00%
386	L 2.093.171.000	28,95%	L 1.141.307.000	...	L 987.896.600	250033,7	1128884	3502486	2,26%	7,00%	2,05%	5,39%
387	L 1.478.670.000	19,53%	L 1.318.090.000	...	L 941.653.500	287191,7	1198019	4180015	2,29%	7,10%	4,05%	4,58%
388	L 1.145.572.000	16,37%	L 1.011.185.000	...	L 797.670.100	221550,9	1245485	3999211	2,20%	6,20%	2,65%	2,47%
389	L 638.404.400	13,03%	L 1.262.743.000	...	L 878.737.900	276763,3	1031903	3374433	3,00%	7,50%	4,64%	4,93%
390	L 824.814.500	11,13%	L 1.208.043.000	...	L 813.829.000	267877,7	1345365	3747030	2,46%	6,96%	4,38%	4,21%
391	L 2.143.037	3,97%	L 1.146.575.000	...	L 828.456.300	2511757	1228387	3468930	2,30%	6,70%	4,13%	4,50%
392	L 956.125.400	15,48%	L 1.058.001.000	...	L 828.456.300	2511757	1204385	3898135	2,89%	7,16%	3,84%	2,08%
393	L 1.502.477.000	22,55%	L 966.441.200	...	L 867.931.000	213501,3	1204385	3531682	2,08%	7,55%	3,63%	3,63%
394	L 1.532.059.000	22,55%	L 1.144.254.000	...	L 760.930.000	213501,3	1242620	3898135	2,89%	7,55%	3,63%	3,63%
395	L 87.9249.000	14,49%	L 1.144.254.000	...	L 760.930.000	213501,3	1242620	3898135	2,89%	7,55%	3,63%	3,63%
396	L 953.035.300	9,12%	L 1.144.254.000	...	L 760.930.000	213501,3	1242620	3898135	2,89%	7,55%	3,63%	3,63%
397	L 427.493.300	5,73%	L 1.105.817.000	...	L 743.143.400	104281,1	104281,1	3217098	2,21%	6,98%	3,27%	3,28%
398	L 427.493.300	5,73%	L 1.105.817.000	...	L 766.943.500	254916,6	1218520	3271371	2,84%	6,22%	4,59%	5,09%
399	L 1.239.619.000	16,53%	L 989.275.900	...	L 811.538.800	221048,3	1208600	3456817	2,38%	6,37%	4,04%	4,20%
400	L 1.688.188.000	24,65%	L 979.268.800	...	L 825.016.200	216461,9	1116742	3468185	2,79%	6,26%	3,38%	2,33%
401	L 1.688.188.000	24,65%	L 979.268.800	...	L 883.484.700	214284,1	1106710	3736183	2,59%	7,17%	4,72%	3,61%

Data-base composto da collezione di dati di simulazione: campionamento LHS, 400 iterazioni (stralcio).

oggetto di analisi critica a partire da alcune precise ragioni di fondo. Prima fra tutte, la presunta indipendenza tra le variabili indipendenti del modello; tale indipendenza, che riferendoci in particolare al contesto immobiliare può essere intesa sul piano fisico, avendo cioè in mente caratteristiche fisiche dei manufatti per loro natura inscindibili, ma anche sul piano psicologico e dei comportamenti dei consumatori (i quali, è stato sottolineato, nell'orientare le proprie scelte applicano anche implicitamente meccanismi di compensazione basati sulla considerazione interrelata delle caratteristiche dei manufatti), è smentita dal mercato immobiliare stesso. Peraltro la regressione, ricordiamo, è tesa a cogliere proprio le diverse manifestazioni di utilità espresse dai consumatori attraverso la disponibilità a pagare, mediante la misurazione di ciascun input in termini di prezzi marginali (Simonotti, 1988; Simonotti, 1989).

I modelli di equazioni strutturali (Bollen, 1989; Corbetta, 1992) - sintesi di analisi fattoriale, modelli causali, modelli di misurazione e reinterpretazione dei modelli causali associati alla Path Analysis - sono in grado di trattare con variabili metriche o assimilabili. Tali modelli, come già detto meno noti in letteratura ma la cui definizione è certamente più sofisticata che non per altri modelli econometrici, superano di fatto i limiti della regressione multipla nel trattamento delle variabili qualitative, introducendo il concetto di "variabile latente"; attraverso le variabili latenti il modello riesce a trattare con variabili mutabili di tipo sconnesso, in maniera più completa che non la varianza o il metodo dei contrasti nella regressione multipla¹⁸.

3.1. Reinterpretazioni della varianza

Le relazioni che legano le medie dei prezzi e le relative varianze osservate, nei mercati, sono un ricorrente riferimento per le analisi. La significatività e l'effettiva utilizzabilità di questo dato rispetto ai mercati finanziari però, come del resto avviene per altri indicatori, risente delle prerogative proprie del mercato immobiliare e anche in questo caso

18) Per una presentazione del modello ed una sua applicazione comparata con la regressione, si veda: Bravi, M.; Fregonara, E., "Structural equations models in real estate appraisal", contributo presentato alla V Conferenza annuale internazionale dell'American Real Estate and Urban Economics Association - A.R.E.U.E.A., 23-25 maggio 1996, Orlando, Florida.

dell'influenza delle componenti di tipo casuale più volte menzionate¹⁹. Come in altri contesti di analisi su problemi immobiliari, il concetto di varianza è riproposto anche in seno all'analisi di rischio costituendone, anzi, un elemento di fondo.

Ancora Hughes osserva che a partire dall'approccio tracciato da Hertz per l'analisi di rischio -il quale, ricordiamo, presuppone la conoscenza delle distribuzioni marginali per tipo di input e la possibilità di produrre un output in termini di probabilità congiunte delle distribuzioni di ingresso (Hertz, 1964)- Pyhrr sviluppa un modello di analisi probabilistica in cui la varianza assume il senso di variabilità dei risultati possibili rispetto ai risultati attesi : *"the measure of risk in this case is the dispersion of returns calculated by the model"*²⁰.

Helton, nel saggio già menzionato in cui tratta alcune tecniche di analisi tra cui la Monte Carlo, ripropone il calcolo del valore medio e della varianza quale possibile via per rappresentare l'incertezza in y (ovvero la sua variabilità); sintetizzando l'esposizione dell'autore, questi due valori possono essere stimati attraverso, rispettivamente, le formule²¹:

$$E(y) = \sum_{i=1}^m y_i / m \quad (4)$$

$$V(y) = \sum_{i=1}^m y_i - E(y)^2 (m - 1) \quad (5)$$

Il termine m indica l'ampiezza del campione generato dalla fase di campionamento (mediante, nel nostro caso, sampling casuale o Latin Hypercube Sampling delle variabili di ingresso incerte (le x_j dell'equa-

19) Come osserva ancora Hughes nell'opera citata, a proposito delle analogie e differenze fra mercati finanziari e mercati immobiliari, la valutazione degli investimenti nel settore finanziario si può avvalere della pratica del calcolo dell'indice Beta; la disponibilità fortemente ridotta di dati sulle transazioni, sugli andamenti della domanda, etc., rendono molto più aleatoria la misurazione del rischio nei contesti immobiliari dove certamente non si può fare riferimento con la stessa facilità a indici di misura media del rischio. Anche la variabilità del ritorno come misura del rischio, proposta da taluni autori, si presenta come procedura inadeguata, data la particolare relazione che lega le dimensioni del rischio ai ritorni attesi e le difficoltà nel reperimento di dati osservati o collezioni di dati storici in base ai quali calcolare le deviazioni standard.

20) Hughes, W.T., *Valuation...*, op. cit., p. 18. Si veda anche Pyhrr, A., A computer simulation model to measure risk in real estate investment, AREUEA Journal, n. 2, 1973.

21) Helton, J.C., "Uncertainty...", op.cit., p. 330. Le espressioni (4) e (5) corrispondono alle (18) e (19) del testo originale.

zione (1)), espresse a mezzo di intervalli di valori e distribuzioni, secondo l'approccio analitico; si ottiene così una sequenza di elementi del campione la cui forma funzionale é:

$$x_i = [x_{i1}, x_{i2}, \dots, x_{in}] , \quad i = 1, 2, \dots, m , \quad (6)$$

con n = numero delle variabili campionate e m = ampiezza del campione. Il termine y_i è dato da una sequenza di risultati della forma:

$$y_i = f(x_{i1}, x_{i2}, \dots, x_{in}) = f(x_i) \quad i = 1, 2, \dots, m . \quad (7)$$

Come afferma Helton, *"in essence, these model evaluations create a mapping from the analysis inputs (i.e. the x_i) to the analysis results (i.e. the y_i) that can be studied in subsequent uncertainty and sensitivity analysis"*²².

3.1.1. Tecniche di riduzione della varianza

Le tecniche di riduzione della varianza (TRV) consistono in procedure finalizzate a ridurre la propagazione dell'incertezza nei risultati del modello (Kleijnen, 1995)²³. Aspetti teorici ed espressioni formali delle stesse sono rintracciabili nella letteratura che tratta i problemi della "verifica e validazione dei modelli", ossia rispettivamente delle attività dirette ad esaminare da un lato il corretto funzionamento del programma di simulazione e, dall'altro, delle attività dirette a comprendere se il modello di simulazione concettuale costituisce una adeguata rappresentazione del sistema che si intende studiare (Law e Kelton, 1991). Riteniamo interessante menzionare le TRV per due ragioni:

- 1) la necessità di oggettivare le analisi basate sulla simulazione, al fine di colmare le eventuali debolezze insite in modelli che operano su una base informativa non osservata;

22) Ibidem. Le espressioni (6) e (7) corrispondono alle (16) e (17) del testo originale.

23) Concetti teorici ed operativi legati a problemi di verifica e validazione dei modelli, in particolare nella ricerca operativa, sono discussi in Kleijnen, J.P.C., Verification and validation of simulation models, in European Journal of Operational Research, n. 82, 1995. Particolare attenzione nel saggio è posta sulle tecniche statistiche, le quali possono fornire dati attendibili circa la qualità dei modelli di simulazione; tali tecniche, specifica l'autore, sono riconducibili a svariate classificazioni.

- 2) l'opportunità di ripensare un concetto -la varianza- ampiamente studiato nelle analisi immobiliari, secondo una declinazione meno consueta.

Data la caratterizzante componente tecnica negli studi di simulazione, è in generale importante verificare il funzionamento dei programmi di calcolo, ma anche -aspetto che più ci compete- la possibilità di "correggere" gli esiti mediante il ricorso ad opportuni stimatori. In letteratura²⁴ è possibile rintracciare la descrizione delle procedure deputate alla verifica dell'output di simulazione, basate sull'ottenimento e verifica dei risultati intermedi di applicazioni simulate ("tracing"), sull'applicazione di test statistici, sulla comparazione tra risultati finali delle simulazioni con output ottenuti per via analitica (prodotta a mezzo di test statistici, a partire da versioni semplificate del programma di simulazione e da soluzioni analitiche note)²⁵.

Le questioni che riguardano i rapporti fra dati del mondo reale e simulazione, ci interessano certamente più da vicino; la validazione -che nel caso della modellizzazione dei sistemi economico-immobiliari è

24) Si veda: Kleijnen, J.P.C., *Verification ...*, op. cit., pp. 146-151. Inoltre, Kleijnen, J.P.C.; Von Groenendaal, W.H., *Simulation...*, op.cit., pp.197-201 ed il Capitolo 11.

25) Nell'esposizione di Kleijnen, nuovamente qui sintetizzata, un passaggio ci pare particolarmente interessante. Partendo dall'ipotesi che l'analista intenda testare le parti del programma deputate alla generazione di campioni a partire da distribuzioni non-uniformi, l'autore consiglia di "stimare la media e varianza della variabile campionata, e di comparare tali statistiche con i valori teorici (...)". Deviazioni sistematiche tra le statistiche osservate ed i valori teorici possono essere indagate attraverso tests parametrici o a distribuzioni libere (...). Casuali (non significative, non sistematiche) deviazioni fra la media del campione e (per dire) y ed il suo valore aspettato y ricorrono sempre (le variabili casuali sono sottolineate). Per ridurre l'effetto di tale deviazione, può essere applicata una tecnica di riduzione della varianza (TRV) detta *control variates*. Questa TRV corregge x , l'output di simulazione (per esempio, tempo medio di attesa), per la deviazione casuale tra la media del campione di input e la media della popolazione:

$$x_c = x + \beta (\mu_y - y)$$

dove una appropriata scelta del coefficiente significa che la varianza del nuovo stimatore x_c è ridotta (...). Invece di testare solo la media o varianza, gli analisti possono testare l'intera distribuzione della variabile casuale. Poi possono applicare un test di goodness-of-fit come i molto noti test del chi-quadro e Kolmogorov-Smirnov (...) (Cfr. Kleijnen, J.P.C., *Verification...*, op.cit., p. 148; traduzione di E. Fregonara).

intesa più nel senso di "predizione" che di "spiegazione" - è orientata, si è accennato, ad indagare "se il modello di simulazione concettuale è una accurata rappresentazione del sistema in studio"²⁶. Considerando quale modello simulato il cash-flow con variabili stocastiche (contrapposto al modello deterministico rappresentato dal quadro dei flussi di cassa tradizionale) la validazione punta a confermare o smentire se il modello costruito è una buona rappresentazione della realtà in oggetto. Per riprendere Klippendorf, tenendo conto della sequenzialità endogena al disegno della ricerca, dunque dell'importanza di individuare, eliminare o ridurre eventuali errori commessi nella "costruzione" del disegno stesso, la validazione utilizza tutto il patrimonio di conoscenza intorno all'oggetto in analisi, al fine di comprendere i legami fra i dati ed il contesto al quale gli stessi si riferiscono, nonchè le modalità con le quali tali legami si esplicano (Klippendorf, 1983).

A parere di chi scrive, alcune delle questioni testè accennate sembrano essere trasferibili all'analisi degli investimenti immobiliari²⁷; purtuttavia, ripercorrere l'esposizione teorica e tecnica (alla quale si rimanda per i dovuti approfondimenti) peraltro esulerebbe dal nostro campo di interesse. Basti comunque accennare che quando, come nel nostro caso, si è costretti ad affrontare la modellizzazione di un sistema economico-immobiliare, è necessario "*explicitly formulate the laws that (...) govern the system under study, which is a system that already exists or is planned to be installed in the real world*". Questo "*implies that the analyst must subjectively decide on the boundary of that system and on the attributes to be quantified in the model*"²⁸. Partendo da questa premessa, una corretta modellizzazione risulta strettamente collegata con l'esatta individuazione delle leggi che influenzano il sistema (gli andamenti dei sovrasistemi economici, le fluttuazioni dei mercati, l'instabilità di alcuni valori, etc.) e della buona misurazione dei dati di ingresso ed uscita del mondo reale - dati che, come noto, nel caso dei sistemi immobiliari si presentano il più delle volte scarsi.

L'influenza delle variazioni temporali nei dati, che come già detto caratterizza il quadro crono-finanziario - si ricollega di fatto al "*time series*

26) Ibidem, p. 151.

27) Si fa riferimento ad alcuni punti trattati a p. 151 e seguenti del saggio citato, sottolineando il fatto che le considerazioni e le esemplificazioni esposte dall'autore sono riferite a contesti diversi dall'economia immobiliare e che, pertanto i parallelismi ed i trasferimenti apportati sono da attribuire all'autore del presente scritto.

28) Ibidem, p. 151.

character of the model inputs and outputs" che, sottolinea l'autore, insieme al "random noise" costituisce aspetto tipico della simulazione²⁹. L'autore illustra alcune tecniche deputate alla comparazione fra serie temporali prodotte quale output di una simulazione e serie storiche temporali prodotte nel mondo reale; tra queste tecniche - applicabili solo nel caso in cui si disponga di dati di input ed output osservabili-, il noto test t di Student opportunamente rivisto (anche in associazione con tecniche di riduzione della varianza)³⁰. Nel caso in cui non si disponga di dati osservabili (ma anche in caso contrario), l'autore consiglia il ricorso alla Sensitivity Analysis³¹; in questo contesto l'obiettivo dell'analisi consiste nella verifica delle conseguenze di cambiamenti nelle previsioni effettuate assumendo, nel nostro caso, una funzione di supporto e di controllo delle componenti previsive dell'analisi dei flussi di cassa³².

29) Ibidem, p.152.

30) Si veda il punto "Some simple techniques for comparing simulated and real data", a p. 152 del saggio citato.

31) Si fa notare che Kleijnen distingue fra Sensitivity Analysis e analisi di sensibilità dell'output rispetto alle distribuzioni di input, condotte nell'analisi di rischio per mezzo del Metodo Monte Carlo; la seconda è definita, più in particolare, Robustness Analysis.

32) Se si procede, come nel caso del software da noi considerato, attraverso l'analisi di regressione, l'espressione formale della relazione input/output, rintracciabile a p. 156 del saggio, risulta:

$$y_i = \beta_0 \sum_{k=1}^K \beta_k x_{ik} \sum_{k=1}^{K-1} \sum_{k'=k+1}^K \beta_{kk'} x_{ik} x_{ik'} e_i$$

dove: y_i = risultato della simulazione nella corsa i degli input di simulazione k , per $i = 1, \dots, n$, con n = numero totale delle corse di simulazione, x_{ik} = valore dell'input di simulazione, k , nella combinazione i , k = effetto di primo ordine dell'input k , kk' = interazione tra gli input k e k' , e_i = errore di approssimazione o di best-fitting nella corsa i . Si tratta in questo caso di un meta-modello di regressione, validato attraverso la comparazione fra output previsto ed output simulato. Per una definizione del concetto di "meta-modello" e per un approfondimento del metodo, si veda Kleijnen, J.P.C.; Von Groenendaal, W.H., Simulation, ..., op. cit., Capitolo 8, in particolare le pagine 149-150. L'espressione -qui riportata per completezza di esposizione- è già stata citata in Curto, R.; Fregonara, E., Decision Tools ..., op. cit., p. 78.

3.2. La regressione in "versione probabilistica"

Abbracciando un campo di applicazione dei metodi multivariati certamente diverso rispetto alle analisi dei fenomeni immobiliari più consuete alla disciplina estimativa, la regressione può essere utilizzata nella risoluzione delle analisi di sensibilità, prodotte nel contesto di studi Monte Carlo (Rubinstein, 1981; Kleijnen e Van Groenendaal, 1992; Helton, 1993).

A questo proposito, il percorso logico tracciato da Helton parte da un modello della forma:

$$y = b_0 + \sum_j b_j x_j \quad (8)$$

il quale viene sviluppato a partire dalla mappizzazione tra gli input dell'analisi ed i risultati della stessa, presentati nell'espressione (7)³³. In quest'ultima, "le x_j sono le variabili di input in considerazione e le b_j sono i coefficienti che devono essere determinati. I coefficienti b_j e altri aspetti della costruzione del modello di regressione mostrati nella (8) possono essere usati per indicare l'importanza delle variabili individuali x_j rispetto all'incertezza in y . La costruzione del modello di regressione in (8) viene considerata per prima. Come mostrato nella (7) esiste una sequenza di valori y_i , con $i = 1, \dots, m$, per la variabile di output. Quando espressa nella forma del modello mostrato nella (8), ciascuna y_i diventa:

$$y_i = b_0 + \sum_j b_j x_{ij} + i \quad i = 1, \dots, m. \quad (9)$$

A questo punto, le b_j sono ancora sconosciute. Si desidera determinare le b_j in qualche modo adatto³⁴.

Il metodo dei minimi quadrati cui molto spesso si ricorre al fine di risolvere le equazioni di regressione, è quindi richiamato dall'autore che ne profila anche gli aspetti formali. In particolare, "per determinare le b_j è conveniente usare la seguente rappresentazione matriciale per le

33) Cfr. Helton, J.C., "Uncertainty ..., op. cit. L'espressione è riportata a p. 349 (formula (62) nel testo originale).

34) Ibidem, p. 349. La traduzione è di E. Fregonara. L'espressione (9) corrisponde alla (63) del testo originale.

eguaglianze in (9):

$$y = Xb + \varepsilon \quad (10)$$

dove

$$y = \begin{bmatrix} y_1 \\ y_m \end{bmatrix}, \quad X = \begin{bmatrix} 1 & x_{11} & \dots & x_{1n} \\ 1 & x_{m1} & \dots & x_{mn} \end{bmatrix}$$

$$b = \begin{bmatrix} b_0 \\ b_n \end{bmatrix} \quad e \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_m \end{bmatrix}$$

Nell'approccio dei minimi quadrati, l'intento è determinare b_j in modo tale che la somma sia minima³⁵.

$$\begin{aligned} S(b) &= \sum_{i=1}^m (y_i - b_0 - \sum_{j=1}^m b_j x_{ij})^2 \\ &= (y - Xb)^T (y - Xb) \end{aligned} \quad (11)$$

L'autore consiglia l'utilizzo della versione Stepwise del modello di regressione per la risoluzione dell'analisi di sensibilità o, in altre parole, dell'analisi dei valori di input rispetto ai risultati, con riferimento all'espressione (8); la Stepwise -che relaziona le x_j alla y considerando le variabili una alla volta- è peraltro il modello di regressione adottato dal software da noi utilizzato³⁶.

35) Ibidem, p. 349. La traduzione è di E. Fregonara. Le espressioni (10) e (11) corrispondono rispettivamente alle (64) e (65) del testo originale.

36) La scelta del modello Stepwise al posto del modello classico con tutte le variabili è giustificata da alcune ragioni:

- del considerevole numero di variabili con le quali si deve trattare, solo un numero più ristretto influisce in maniera significativa sulla variazione dell'output, mentre il contributo all'incertezza di molte voci è pressochè trascurabile;
- la presenza di fenomeni di collinearità tra le variabili provoca una certa instabilità nei coefficienti di regressione, particolarmente sensibili alle variazioni eventuali ap-

Come è noto, il percorso procedurale della Stepwise Regression Analysis è basato sulla costruzione di una sequenza di modelli di regressione, aggiungendo una variabile alla volta secondo un ordine di importanza delle stesse stabilito in precedenza (attraverso p.e. il calcolo dei valori di R^2 o dei valori assoluti dei coefficienti di regressione standardizzati). La costruzione di nuovi modelli si arresta quando l'aggiunta di una nuova variabile provoca un incremento scarsamente significativo nella variazione della variabile dipendente; l'opportunità di fermare il processo di costruzione di modelli successivi, oltre che di aggiungere o cancellare variabili ai modelli di regressione, può poi essere verificata per mezzo del calcolo di test specifici, quali il *t-test* (che esprime la significatività di ciascun coefficiente di regressione), l' *value*, l'indice *PRESS* (*Predicted Error Sum of Squares*).

Trasferendo, con i dovuti adattamenti e semplificazioni rispetto al nostro contesto di applicazione, l'esposizione dell'autore e traendo spunto da un esempio dallo stesso esposto³⁷, possiamo per ipotesi considerare la variabile di output VAN quale dipendente nell'equazione di regressione: i coefficienti marginali calcolati per mezzo della regressione Stepwise per ciascuna variabile di input misurano la sensibilità dell'output rispetto alla particolare distribuzione di ingresso selezionata. Se per esempio consideriamo un numero $n = 10$ di variabili indipendenti (quali Costo di costruzione, Prezzo di vendita, Tasso di interesse, etc.) per un campione di dimensione $m = 30$, ottenuto con l'applicazione di un opportuno metodo di campionamento, si ha:

$$[y_i, x_{i1}, x_{i2}, \dots, x_{i10}], \text{ con } i = 1, \dots, 30, \quad (12)$$

dove y_i è il valore dell'output (esprimibile anche in forma grafica attraverso la funzione di distribuzione del VAN), x_{in} sono i valori per le variabili ritenute critiche della Tabella ACR per l'*i*-esimo elemento del campione.

portate al modello di regressione (instabilità difficilmente coglibile da un modello regressivo contenente tutte le variabili);

- possibilità indesiderata che si verifichino fenomeni di overfitting dei dati, fenomeno che "avviene quando il modello di regressione tenta di incontrare le predizioni associate ad elementi singoli del campione piuttosto che incontrare gli andamenti mostrati dagli elementi del campione collettivamente" (Helton, op.cit. p. 353).

37) Ibidem, pp. 354-356.

L'applicazione è preceduta dal calcolo della matrice di correlazione -in questo caso di ordine $(10+1)*(10+1)$ - tra il VAN e ciascuna delle variabili di input, in modo da consentire l'ordinamento di importanza e, dunque, l'ordine di entrata delle stesse nei modelli. Se il primo passo seleziona dalla matrice di correlazione la variabile di input x_j , che supponiamo sia "Costo di costruzione", ossia la variabile che mostra in valore assoluto la più elevata correlazione con il VAN, il modello di regressione che ne risulta è:

$$\hat{y} = b_0 + (\text{coefficiente di correlazione}) * (\text{Costo di costruzione})_{\sim},$$

dal quale possono essere calcolati il valore di R^2 , di \hat{y} , il coefficiente di regressione standardizzato, etc. Si procede quindi con il calcolo di b , secondo l'approccio dei minimi quadrati, consentito dal fatto che si dispone di un numero di elementi del campione che supera il numero di variabili indipendenti³⁸. La costruzione del secondo modello parte dalla selezione di una seconda variabile di input x_j , per esempio "Prezzo di vendita", la quale secondo una logica analoga alla precedente presenta il più elevato impatto sulle variazioni del VAN non spiegate dalla variabile del primo modello. Tale variabile può essere espressa dalla forma funzionale:

$$\tilde{y} = y - \hat{y} = y - (b_0 + \text{coefficiente di correlazione} * \text{Costo di costruzione})$$

dove \hat{y} è definita nella precedente equazione e calcolando le correlazioni fra y e le rimanenti variabili. Il secondo modello assume dunque la forma:

$$\hat{y} = b_0 + (\text{coefficiente di correlazione}) * (\text{Costo di costruzione}) + (\text{coefficiente di correlazione}) * (\text{Prezzo di vendita}),$$

dove $n = 2$ e $m = 30$. Analogamente, il terzo modello sarà costruito includendo quella variabile che mostrerà più elevato impatto sulla variazione del VAN (per esempio la variabile Tasso di interesse) non spiegata dalle due precedenti. La nuova variabile avrà la forma:

$$\tilde{y} = y - \hat{y} = y - (b_0 + \text{coefficiente di correlazione} * \text{Costo di costruzione} - \text{coefficiente di correlazione} * \text{Prezzo di vendita})$$

38) Ibidem. L'approccio è formalizzato alle pagine 349-350 del saggio.

dove \hat{y} è definito nel modello precedente. Il terzo modello assumerà dunque la forma:

$$\hat{y} = b_0 + (\text{coefficiente di correlazione}) * (\text{Costo di costruzione}) + (\text{coefficiente di correlazione}) * (\text{Prezzo di vendita}) + (\text{coefficiente di correlazione}) * (\text{Tasso di interesse}).$$

La procedura prosegue per un certo numero di modelli; l'individuazione del punto di arresto del processo, si è detto, è un passaggio delicato reso risolvibile dal calcolo di alcuni indicatori specifici, quali l'indice R^2 , l'indice PRESS, il confronto fra gli R^2 .³⁹ La sostanza del metodo Stepwise consiste a questo punto nel confronto fra i valori dei coefficienti di regressione e degli indici R^2 calcolati per i diversi modelli, per esempio per sottrazione, ottenendo così le variazioni nell'output dovute all'aggiunta di una nuova variabile: se la variabile Costo di costruzione spiega il 50% della variabilità del VAN, mentre Costo di costruzione + Prezzo di vendita spiegano il 70%, la variabile Prezzo di vendita da sola spiega circa il 20% della variazione.

L'esemplificazione appena fornita del metodo Stepwise è senza dubbio scontata, data la diffusione del metodo stesso; tuttavia, ci è parsa utile dal momento che su tale procedura multivariata, particolarmente adatta a calcolare i valori di regressione in presenza di valori di input multipli (poichè, si ricorda, le variabili che non forniscono un contributo significativo sono rimosse dal modello), si imposta il percorso operativo proposto dal software @Risk; più nello specifico, quest'ultimo si fonda su due tecniche analitiche:

- 1) l'analisi di regressione. Attraverso la regressione Stepwise, il software fornisce i valori di sensibilità nella forma di variazioni normalizzate dei coefficienti di regressione⁴⁰: nell'output "*sensitivity report*", sono

39) In generale si può considerare che il modello debba essere concluso quando il valore dell'indice non risulta inferiore ad un indice prefissato (0.02 in alcuni studi), quando cioè una nuova variabile non produce più variazione significativa sull'output. L'indice informa inoltre, in ciascuno stadio, sulla significatività di una variabile già selezionata per un precedente modello (se la variabile non è più significativa eccede un livello specificato, 0.05 in alcune analisi). Quando una variabile risulta superflua, viene tolta dal modello, ma può essere reinserita in uno successivo, qualora sia opportuno.

40) Le considerazioni teoriche e tecniche sulla Multivariate Stepwise Regression a sulla Rank Order Correlation, secondo l'approccio seguito nel programma, sono una sintesi opportunamente integrata di quanto riportato in: Palisade Corp., The @Risk Troubleshooting Guide, Palisade Corporation, Newfield, NY 1995, sezione 6, pp. 66-67.

riportati i coefficienti di regressione normalizzati relativi a ciascuna voce di ingresso analizzata. Negli studi più tradizionali ai contesti immobiliari, si menziona, l'indice di determinazione R^2 esprime il grado complessivo di accostamento del modello ai dati osservati: valori pari ad 1 (o -1) indicano il caso di perfetta sovrapposizione della retta interpolante ai dati rilevati, mentre valori pari a 0 indicano il caso di assenza di accostamento. Tale indice, espresso dal rapporto fra variabilità spiegata e variabilità totale, rappresenta la percentuale secondo la quale una caratteristica è in grado di spiegare la variabile dipendente (il prezzo), e risulta normalmente accettabile per valori inferiori a 0,9/0,95; la parte residua rappresenta la percentuale di prezzo che non è spiegata dal modello. Nel nostro contesto di analisi, il coefficiente di regressione multipla esprime l'entità dell'influenza di ciascun parametro sul valore dell'output: valori di regressione pari a 1 (0 -1) indicano una perfetta correlazione tra due variabili, mentre un valore di regressione pari a 0 indica l'assenza di relazione significativa fra le stesse. Il software, in particolare, per valori di R^2 inferiori a circa il 60%, considera la regressione lineare non in grado di spiegare la relazione in oggetto e, in questo caso, consiglia il ricorso ad un altro metodo di analisi;

- 2) il calcolo del grado di correlazione, espresso mediante il computo dei coefficienti di correlazione tra i valori di output e ciascun insieme di valori di input campionati: è calcolata nel nostro caso la correlazione fra il VAN ed i campioni per ciascuna distribuzione di input. Questo passaggio è risolto applicando la Rank Correlation Analysis, che prevede il calcolo del rango dei coefficienti di correlazione di Spearman. Il software fornisce valori che possono variare fra -1 e 1: ricordando che " la correlazione è una misura quantitativa della forza di una relazione fra due variabili", nel caso della correlazione lineare i valori estremi indicano rispettivamente correlazione completa negativa e positiva fra due variabili, il valore 0 indica l'assenza di correlazione (le variabili sono indipendenti); valori intermedi indicano correlazioni parziali. Rispetto al contesto simulativo, inoltre, la forza della relazione fra due variabili si riscontra nella modalità con cui avviene il campionamento: nel caso di correlazione completa positiva fra due variabili, infatti, quando viene campionato un valore di input per esempio elevato il valore di output verrà campionato elevato; nel caso inverso, ossia quando il valore di input viene campionato alto, il

valore di output verrà campionato basso⁴¹. Sulla base di questa logica, la rank order correlation analysis punta ad individuare le relazioni fra due insiemi di dati attraverso la comparazione del rango di ciascun valore all'interno di un insieme di dati⁴². Certamente l'indicatore R2 è a noi più familiare; però, data la natura delle variabili con cui si deve trattare, definite per mezzo delle proprie funzioni di distribuzione di probabilità, non siamo sempre in presenza di correlazione di tipo lineare. In questo caso, oltre che qualora si presentino valori di R2 bassi (che dimostrano appunto la presenza di relazione non lineare), la rank order correlation risulta più adatta; inoltre, la *rank correlation* ha il vantaggio di essere adattabile a dati di tipo qualitativo (Pellerey, 1998)⁴³.

Un'ultima riflessione riguarda infine il fenomeno della multicollinearità, che sappiamo essere un grande limite della regressione multipla e che, come tale, è stato oggetto di trattazione in letteratura.

Come si riscontra negli studi immobiliari classici, l'effetto della multicollinearità è provocato fondamentalmente dalla presenza di caratteristiche indivisibili dei beni in oggetto di valutazione (rispetto a variabili intrinseche, estrinseche e territoriali); tali effetti vengono combattuti in diversi modi, primo fra tutti il testaggio dei modelli al fine di individuare ed includere nel modello finale solo le variabili in grado di rappresentare caratteristiche rilevanti sul prezzo. Menzionando che si presentano fenomeni di multicollinearità quando si trasgredisce la condizione fondamentale dei modelli lineari, ossia l'indipendenza tra le variabili -indipendenti, appunto- del modello, sarebbe auspicabile che, nel nostro caso, ciascuna delle voci sia in condizione di propagare incertezza sull'output indipendentemente dalla altre. La presenza di multicollinearità, peraltro, può essere avvertita allorquando l'output della sensitivity analysis riproduce, rispetto a due variabili, risultati contraddittori nell'analisi di regressione e nella rank order correlation analysis (in tal caso, una variabile dovrebbe essere rimossa).

41) Ibidem, sezione 6, p. 67.

42) Ibidem. Nel testo si specifica che il rango viene attribuito dapprima ordinando i dati dal più basso al più alto, ed assegnando poi ad ogni livello un numero (rango), che riflette la posizione ordinale dei dati stessi.

43) Si veda l'osservazione 7.1, a p. 178 del testo: Pellerey, F., Elementi di statistica per le applicazioni, Celid, Torino 1998.

3.3. Modelli di equazioni strutturali: nuovi impieghi possibili?

Molte applicazioni pratiche hanno evidenziato come la regressione multipla non può essere considerata quale migliore tecnica per la valutazione immobiliare, spingendo ad interrogarsi sull'esistenza di modelli alternativi migliori. Con questo presupposto e per questo scopo, si è posta attenzione verso i modelli di equazioni strutturali (MES).

In conclusione di questo lavoro, accenniamo alle potenzialità dei MES, con presupposti e finalità diverse da quelle dei precedenti studi: da un lato, poniamo quale obiettivo la quantificazione del rischio e, dall'altro, l'applicazione di tecniche per la riduzione della varianza.

Il primo punto si basa sulla precisa convinzione che il rischio può essere considerato quale variabile latente misurata a mezzo di variabili osservabili. Il secondo punto parte invece da un interrogativo: interpretando in chiave estimativa quanto detto nel Sottoparagrafo 3.1.1, possiamo intendere la riduzione della propagazione dell'incertezza nei termini di "ripulitura" della parte di errore dalla fascia di varianza dell'output al fine di ottenere la fascia "pura" di rischio, ossia nei termini di pratica per la "correzione degli stimatori del modello"? Se questa affermazione è concettualmente valida, i modelli di equazioni strutturali possono essere assimilati a tecniche per la riduzione della varianza o, secondo un'altra sfumatura, a procedure per la verifica del modello di simulazione.

Come si è già detto (Paragrafo 3), la statistica descrittiva sulla quale si giustifica l'analisi della varianza non consente a quest'ultima di risolvere il problema dell'interazione fra le variabili in termini previsivi. Questo limite è in qualche modo superato dall'analisi di regressione - procedimento pluriparametrico in grado, ricordiamo, di spiegare le relazioni di tipo causa-effetto esistenti fra le variabili quantitative, sulla base del principio della interpolazione fra dati osservati della funzione teorica che meglio approssima i dati stessi-. La regressione, abbiamo già detto, pone però il problema del trattamento delle variabili qualitative. Tale limite, assieme peraltro ad altre difficoltà poste per esempio dalla non separabilità di certe variabili esplicative del fenomeno in studio, dalla presenza di correlazione o di relazioni non solo pure fra le variabili stesse, è superato nel MES. Questo introduce l'aspetto della "causalità" (Corbetta, 1992) nelle fasi di analisi preposte alla modellizzazione dei legami -causali appunto- fra le variabili. Tale fatto conferisce al MES maggiore potenzialità, dal momento che il modello di regressione esprime il legame con le variabili indipendenti in termini di correlazione più

che di causalità⁴⁴. La difficoltà nel trattamento della causalità-rappresentativa peraltro di un concetto di natura teorica, fatto non estraneo alla sfera dei fenomeni economici- è insita nella sua non immediata riconoscibilità.

Inoltre, il limite della regressione nel trattamento delle variabili qualitative è superato nel MES dall'introduzione delle variabili latenti, concetti teorici derivati dall'analisi fattoriale.

Un altro punto, particolarmente interessante poiché più contestuale ai temi trattati nel corso di questo scritto, riguarda il problema della verifica del modello, chiamando in causa il "principio della falsificazione". Nel caso della regressione multipla la verifica del modello stimato -ossia la verifica della bontà dell'adattamento teorico ai dati osservati- viene prodotta attraverso l' R^2 . La verifica della capacità inferenziale dei modelli, in altre parole l'affidabilità statistica dei coefficienti di regressione calcolati, passa attraverso la "prova delle ipotesi": operativamente, il calcolo dell'errore (attraverso l'indice dell'errore standard) permette di stabilire il grado di fiducia che possiamo riporre sulle stime prodotte, confrontando ed analizzando il distacco (scarto) fra dati osservati e dati prodotti dal modello teorico. I test prodotti in regressione, però, si limitano a constatare la non erroneità del modello costruito, ma non si esprimono sul fatto che il modello in esame sia "quello giusto". I passaggi di verifica del modello, nel MES, costituiscono perciò un ulteriore superamento della regressione multipla.

Infine, il MES rispetto alla regressione prevede una più semplice ed approfondita modalità per misurare l'errore associato con le variabili, dipendenti ed indipendenti.

Siamo consapevoli delle differenze che sussistono fra contesti di valutazione estimativi (nei quali più tradizionalmente si collocano le stime pluriparametriche) ed i contesti dei giudizi di convenienza economica (nei quali si colloca l'ACR). Purtroppo, come abbiamo in precedenza visto a proposito della regressione multipla, la sfera delle procedure econometriche può efficacemente essere integrata in precisi momenti di analisi di un quadro crono-finanziario. Il MES affronta il problema della misura delle variabili e della scelta della forma funzionale più adatta a rappresentarle ed a riprodurre le relazioni fra le stesse (problema inso-

44) Per la presentazione di un caso-studio orientato al confronto fra i risultati ottenuti dall'applicazione del modello di regressione e del MES al medesimo campione di analisi, si rimanda al saggio Bravi, M.; Fregonara, E., *Structural Equations ...*, op. cit.

luto dalla regressione). Se la modellizzazione è prodotta a fini estimativi, il problema interessa il trattamento delle variabili rappresentative, per esempio, delle caratteristiche di beni disomogenei quali tipicamente i beni immobiliari. Se invece ci spostiamo verso la valutazione di una ipotesi di investimento modellizzata dall'ACR probabilistica, il problema si pone nei termini, illustrati in precedenza, di una analisi multivariata: si propone dunque l'utilizzo dei sistemi di equazioni strutturali per la risoluzione dell'analisi di sensibilità.

Questo aspetto già costituisce uno spunto di ricerca da sviluppare e sperimentare sul piano empirico (essendo un elemento di novità almeno per quanto riguarda le applicazioni da noi finora prodotte); un più attento esame del modello induce a soffermare l'attenzione su un altro punto.

Il passaggio ai metodi di stima probalistici (e, nel nostro caso, ai metodi per la formulazione dei giudizi di convenienza economica), si è visto negli studi, introduce il problema della misurazione dell'errore della stima stessa; operando, in linea con i nostri ragionamenti, su campioni di dati simulati, si presenta anche il problema della individuazione dell'errore dovuto alla particolare origine dei dati in oggetto: "errore" che, nella letteratura sulla verifica e validazione dei modelli, potrebbe essere ricondotto alla pratica della simulazione. Il MES, qualora sia applicato per fini analitici-estimativi su un campionamento da simulazione, opera una analisi alternativa al metodo tradizionale dei minimi quadrati ordinari finalizzata ad individuare i "contributi marginali impliciti" di ciascuna variabile stocastica al raggiungimento di un certo output e, sempre in analogia con il sistema di equazioni strutturali classico, è in grado di misurare l'errore associato alle variabili dipendenti ed indipendenti del modello. Si sottolinea che, proprio in virtù di una prerogativa dei modelli strutturali, l'errore viene considerato distinguendo una componente stocastica ed una componente di misurazione: il modello distingue -sulla base dei criteri di attendibilità e di validità della stima- l'errore dovuto alla mancanza di stabilità commessa nella misurazione (in particolare rispetto a misurazioni ripetute nel tempo) e l'errore invece "sistematico", connesso con la capacità del modello di misurazione di produrre le misurazioni necessarie (aspetto che si riallaccia al problema, precedentemente richiamato, di "validazione" del modello). Per mezzo del calcolo del coefficiente di attendibilità, è possibile correggere la varianza della variabile dipendente (osservata), la quale viene depurata della parte di errore stocastico

(non sistematico). Tale aspetto presenta senza dubbio dei legami con la *ridge regression* che, ricordiamo, è in grado di agire rispetto all'errore degli stimatori: mentre la *ridge*, però, riduce la varianza dello stimatore, il MES attribuisce una parte della varianza all'errore di misurazione e, inoltre, quantifica il peso di tale componente di errore nelle relazioni fra le singole variabili.

Alla luce di queste ultime considerazioni certamente pare rafforzata l'ipotesi di considerare il MES per la riduzione della varianza, nel senso in precedenza tracciato; senza dubbio, se pensiamo di dover operare misurazioni ripetute nel tempo -come l'ACR richiede- la possibilità di formulare quantificazioni delle quote di rischio più attendibili, riducendo l'incertezza naturalmente associata con analisi a sfondo previsivo, risulta quantomeno da esplorare.

Riferimenti bibliografici

Bravi, M.; Fregonara, E., *Structural equations models in real estate appraisal*, contributo presentato alla V Conferenza annuale internazionale dell'American Real Estate and Urban Economics Association - A.R.E.U.E.A., 23-25 maggio 1996, Orlando, Florida.

Bollen, K.A., *Structural equations with latent variables*, Wiley and Sons, New York 1989.

Brealy, R.A.; Myers, S.C., *Principi di finanza aziendale*, Mc Graw-Hill Italia, Milano 1993.

Corbetta, P., *Metodi di analisi multivariata per le scienze sociali*, Il Mulino, Bologna 1992.

Curto, R., *Lo stato dell'arte nell'ambito dell'estimo urbano*, in Atti del XXV Incontro Ce.S.E.T., Roma, 2-3 Ottobre 1995.

Curto, R.; Fregonara E., *Il controllo del rischio e dell'incertezza negli investimenti immobiliari: la probability analysis*, in Genio Rurale, n. 9, 1997.

Curto, R.; Fregonara, E., *Decision tools for investments in the real estate sector with risk and uncertainty elements*, in Jahrbuch fuer Regionalwissenschaft, n.19, 1999.

Fregonara, E., *Forecasting and control of market risk in real estate investments*, in Proceedings della IX Conferenza annuale SRA-E, Society of Risk Analysis - Europe, Rotterdam, ottobre 1999.

Kleijnen, J.P.C.; Von Groenendaal, W.H., *Simulation. A Statistical Perspective*, Wiley, New York 1992.

Kleijnen, J.P.C., *Verification and validation of simulation models*, in European Journal of Operational Research, n. 82, 1995.

Klippendorf, K., *Analisi del contenuto. Introduzione metodologica*, ERI, Torino 1983.

Helton, J.C., *Uncertainty and sensitivity analysis techniques for use in performance assessment for radioactive waste disposal*, in Reliability Engineering and System Safety, n. 42, 1993.

Hertz, D., *Risk analysis in capital investment*, Harvard Business Review, n. 42, 1964.

Hughes, W.T., *Valuation and risk analysis of lease properties through simulation: an objective Monte Carlo technique*, dissertazione di dottorato, Università della Georgia, 1990.

Law, A.M.; Kelton, W.D., *Simulation modeling and analysis*, Mc Graw-Hill, 1991.

Palisade Corp., *@Risk, Advanced risk analysis for spreadsheets*, User's Guide, Palisade Corporation, Newfield, NY 1995.

Palisade Corp., *The @Risk Troubleshooting Guide*, Palisade Corporation, Newfield, NY 1995.

Pellerey, F., *Elementi di statistica per le applicazioni*, Celid, Torino 1998.

Prizzon, F., *Gli investimenti immobiliari*, Celid, Torino 1995.

Rubinstein, R.Y., *Simulation and the Monte Carlo Method*, Wiley, 1981.

Simonotti, M., *L'analisi di regressione nelle valutazioni immobiliari*, in Studi di economia e diritto, n. 3, 1988.

Simonotti, M., *Fondamenti di metodologia estimativa*, Liguori, Napoli 1989.

Stephens, M.E.; Goodwin, B.W.; Andres, T.H., *Deriving parameter probability density functions*, in Reliability Engineering and System Safety, 42, 1993.